DOCUMENT RESUME

ED 341 728 TM 017 923

AUTHOR	Mittag, Kathleen Cage
TITLE	Correcting for Systematic Bias in Sample Estimates of Population Variances: Why Do We Divide by n-1?
PUB DATE	Jan 92
NOTE	31p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January-February 1992).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	College Mathematics; Computer Software; *Equations (Mathematics); *Estimation (Mathematics); Higher Education; *Mathematical Models; Methods Courses; Monte Carlo Methods; Research Methodology; Research Needs; *Sampling; *Statistical Bias; Statistics
IDENTIFIERS	*Population Parameters; *Variance (Statistical)

ABSTRACT

An important topic presented in introductory statistics courses is the estimation of population parameters using samples. Students learn that when estimating population variances using sample data, we always get an underestimate of the population variance if we divide by n rather than n-1. One implication of this correction is that the degree of bias gets smaller as the sample gets larger and larger. This paper explains the nature of bias and correction in the estimated variance and discusses the properties of a good estimator (unbiasedness, consistency, efficiency, and sufficiency). A BASIC computer program that is based on Monte Carlo methods is introduced, which can be used to teach students the concept of bias in estimating variance. The program is included in this paper. This type of treatment is needed because surprisingly few students or researchers understand this bias and why a correction for bias is needed. One table and three graphs summarize the analyses. A 10-item list of references is included, and two appendices present the computer program and five examples of its use. (Author/SLD)

Correcting for Systematic Bias in Sample Estimates of Population Variances: Why Do We Divide by n -1?

Kathleen Cage Mittag
Texas A&M University 77843-4232

Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX, January 31, 1992.

ABSTRACT

An important topic presented in introductory statistics courses is the estimation of population parameters using samples. Students learn that when estimating population variances using sample data, we always get an underestimate of the population variance if we divide by n rather than by n-1. One implication of this correction is that the degree of bias gets smaller as the sa.nple gets larger and larger. This paper explains the nature of bias and correction in the estimated variance and discusses the properties of a "good" estimator. A computer program was included to illustrate the bias concept and is included in this paper. This type of treatment is needed, because surprising few students or researchers understand this bias and why a correction for bias is needed.

An important topic presented in introductory statistics courses is the estimation of population parameters using samples. In many statistical studies it may be too costly, too time-consuming, or simply impossible to gather data from the entire population. Methods have been developed to estimate these population parameters and this paper will explain the nature of bias and correction in the estimated variance and discuss the properties of a "good" estimator. A BASIC computer program for an IBM PC is included in Appendix 1; this program can be used to teach students the concept of bias in estimating variance.

Variance is what is called a point estimate. A point estimate is computed from a given sample and has a single numerical value that acts as an approximation of the population parameter. Interval estimates (not discussed in this paper) specify limits between which population parameters fall with a given probability. These interval estimates are called confidence intervals.

Before discussing parameter estimates, a review of the basic computational statistics for population mean, variance, and standard deviation will be presented (Harnett, 1970; Ott, 1988). The mean, a measure of central tendency, is defined as:

$$\mu = E[\underline{x}] = (1/\underline{N}) \sum_{\underline{i}=1}^{\underline{n}} \underline{x_i} \text{ , where } \underline{N} = \text{number in population} \tag{1}$$

The variance, a measure of variability, is defined as:

$$\sigma^2 = E[(\underline{x} - \mu)^2] = (1/\underline{N}) \Sigma (\underline{x} - \mu)^2 \tag{2}$$

The standard deviation, a measure of variability, is defined as the square root of the variance. It is used because the variance is in a squared metric, and people are more comfortable thinking in units of dollars rather than squared dollars, or IQ rather than squared IQ, and so forth.

$$\sigma = \sqrt{(\sigma^2)} \tag{3}$$

Since population data are seldom available, it is often necessary to estimate parameters using sample data. There are four criteria which are considered when deciding if an estimator is a "good" estimator. These criteria are unbiasedness, consistency, efficiency, and sufficiency (Harnett, 1970; Khazanie, 1990). The cost of making incorrect estimates from sample data should be minimized; therefore, it is very important to chose the correct estimation procedure. In this paper, parameters will be referred to as θ. Parameter estimates will be referred to as θ. An example is if $\theta = \sigma^2$, then $\theta = s^2$.

UNBIASEDNESS

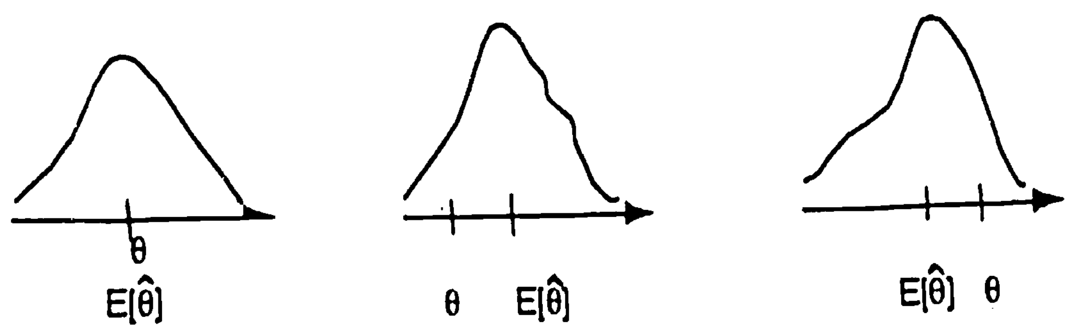
Unbiasedness is the first property of a "good" estimator. Carl Gauss is given credit for first presenting this concept. Unbiasedness is defined by Harnett (1970, p. 188) as the following:

> An estimator is said to be unbiased if the expected value of the estimator is equal to the parameter being estimated, or if
>
> $$E[\theta] = \theta \tag{4}$$

Ideally, the bias should be equal zero. A biased estimator will either underestimate (Figure 1(c)) or overestimate (Figure 1(b)) the parameter θ. If an estimator is "good" and several samples are taken from a population, then the mean value of these samples should be close to the parameter value (Figure 1(a)). Khazanie (1990) illustrated these unbiasedness concepts as follows:

FIGURE 1

Curves represent sampling distributions of θ



(a). $\hat{\theta}$ is unbiased estimator of θ

(b). $\hat{\theta}$ overestimates θ

(c). $\hat{\theta}$ underestimates θ

The sample mean (M), the most widely used estimator, is an unbiased estimator of μ. This fact can be shown as follows (Harnett, 1970, p. 159).

$$\text{Define } \underline{M} = (1/\underline{n})\,(\underline{x}_1 + \underline{x}_2 + \ldots + \underline{x}_n) \tag{5}$$

$$\begin{aligned} E[\underline{M}] &= E[(1/\underline{n})\,(\underline{x}_1 + \underline{x}_2 + \ldots + \underline{x}_n)] \\ &= 1/\underline{n}\, E[\,\underline{x}_1 + \underline{x}_2 + \ldots + \underline{x}_n] \\ &= 1/\underline{n}\,(\,E[\underline{x}_1] + E[\underline{x}_2] + \ldots + E[\underline{x}_n]) \\ &= 1/\underline{n}\,(\,\mu + \mu + \ldots + \mu \end{aligned}$$

$= 1/\underline{n}\ (\underline{n}\mu)$

$\mu_{\underline{M}} \quad = \mu$

$\therefore\ E[\underline{M}] = \mu$ (6)

It would be nice if the sample variance ($s^2 = 1/\underline{n}\ [\Sigma\ (\underline{x}-\underline{M})^2]$), the second most widely used estimator, was also an unbiased estimator of the population variance (σ^2), but s^2 is **not** an unbiased estimator of σ^2 ($E[s^2] \neq \sigma^2$) (Harnett, 1970). It is a fact that s^2 always underestimates σ^2 by a factor of $(\underline{n}-1)/\underline{n}$. The following relationship results from that fact:

$E[s^2] = \sigma^2\{(\underline{n}-1)/\underline{n}\}$ (7)

or by rewriting (7)

$E[s^2] = \sigma^2 - \sigma^2/\underline{n}$ (8)

From formula (8) it can be seen that the bias is equal $\sigma^2/\underline{n}$. If $\underline{n}$ is large, then σ^2/n becomes very small. That fact reinforces the idea that it is important to have a large sample size, if possible. The value of $\sigma^2/\underline{n}$ can be important, as illustrated in the next example. Assume that the population variance is σ^2 = 50 and calculate the estimate variance, s^2, from samples of size $\underline{n}$=5, $\underline{n}$=10, and $\underline{n}$=20. The estimate from $\underline{n}$=5 will be 20% too low, since

$E[s^2] = 50 - (50/5) = 40.$

The estimate from $\underline{n}$=10 would be 10% too low and the estimate from $\underline{n}$=20 would be 5% too low. This illustrates how sample size effects the underestimates of variance.

It is very easy to correct for this bias in the variance formula (7). All that needs to be done is to multiply formula (7) on both sides by the reciprocal of $(\underline{n}-1)/\underline{n}$, which would be $\underline{n}/(\underline{n}-1)$ (Harnett, 1970).

$$\{\underline{n}/(\underline{n}-1)\}\, E[s^2] = \{\underline{n}/(\underline{n}-1)\}\{(1/\underline{n})\, \Sigma\, (\underline{x}-\underline{M})^2\}$$

$$S^2 = \{1/(\underline{n}-1)\}\, \Sigma\, (\underline{x}-\underline{M})^2 \tag{9}$$

The formula (9) will be referred to as the **unbiased estimate of σ^2** and denoted by S^2.

Formula (7) is essential in deriving the unbiased variance estimate; ther... re, the proof of formula (7) in included in this paper. The proof is as fol'ows (Harnett, 1970):

In the first step of the proof, $(\underline{x}-\mu) - (\underline{M}-\mu)$ is substituted for the term $(\underline{x}-\underline{M})$, since they are equivalent mathematically.

$$E[s^2] = E[(1/\underline{n})\, \Sigma\, (\underline{x}-\underline{M})^2]$$

$$= (1/\underline{n})\, E[\Sigma\{(\underline{x}-\mu) - (\underline{M}-\mu)\}^2]$$

Note: Since $(\underline{a}+\underline{b})^2 = \underline{a}^2 + 2\underline{ab} + \underline{b}^2$, the next step follows.

$$= (1/\underline{n})\, E[\Sigma(\underline{x}-\mu)^2] - (2/\underline{n})\, E[\Sigma(\underline{M}-\mu)(\underline{x}-\mu)] + (1/\underline{n})\, E[\Sigma(\underline{M}-\mu)^2]$$

Note: In the second term, $\Sigma(\underline{M}-\mu)$ is a constant, so it can be taken outside the expectation sign.

$$= (1/\underline{n})\, E[\Sigma(\underline{x}-\mu)^2] - (2/\underline{n})\, \Sigma(\underline{M}-\mu)\, E[(\underline{x}-\mu)] + (1/\underline{n})\, E[\Sigma(\underline{M}-\mu)^2]$$

Note: Let $E[(\underline{x}-\mu)] = (\underline{M}-\mu)$

Let $E[(\underline{x}-\mu)^2] = \sigma^2$

Let $E[(\underline{M}-\mu)^2] = \sigma^2/\underline{n}$

$= (1/\underline{n}) \Sigma\sigma^2 - (2/\underline{n}) \Sigma(\underline{M}-\mu)(\underline{M}-\mu) + (1/\underline{n}) \Sigma(\sigma^2/\underline{n})$

$= (1/\underline{n}) \Sigma\sigma^2 - (2/\underline{n}) \Sigma(\underline{M}-\mu)^2 + (1/\underline{n}) \Sigma(\sigma^2/\underline{n})$

$= \sigma^2 - 2(\sigma^2/\underline{n}) + (\sigma^2/\underline{n})$

$= \sigma^2(\underline{n}/\underline{n} - 2/\underline{n} + 1/\underline{n})$

$= \sigma^2\{(\underline{n}-1)/\underline{n}\}$

$\therefore\ E[s^2] = \sigma^2\{(\underline{n}-1)/\underline{n}\}$

Unbiasedness has one weakness in that it requires only the average value of $\hat{\theta}$ equal θ. The values of $\hat{\theta}$ can be very far from θ and still average θ. The next property, consistency, takes the variability of $\hat{\theta}$ into consideration.

CONSISTENCY

The definition and properties of consistency given by Harnett (1970, p. 191) are:

Definition: An estimator is said to be consistent if it yields estimates which approach the population parameter being estimated as $\underline{n}$ becomes larger.

Properties: 1). $\mathrm{Var}\,(\hat{\theta}) \to 0$ as $\underline{n} \to \infty$

2). $\hat{\theta}$ is unbiased ($E[\hat{\theta}] = \theta$)

Rahman (1968, p. 301) emphasized that "estimates" can be both consistent and unbiased, neither, or one or the other in the following quotation:

> Nevertheless, it is to be emphasized that consistency is a very different concept from unbiasedness, and it is also derived from a different theory of estimation. (Unbiasedness is derived from the theory of least squares.) As such, a consistent estimate may or may not be unbiased. Conversely, an unbiased estimate may or may not be consistent. Despite this, there exist estimates (such as the sample mean) which are both unbiased and consistent.

R. A. Fisher introduced the consistency property in the 1920's.

EFFICIENCY

The third property, efficiency, concerns the reliability of the estimate of θ for a given sample size. Khazanie (1990, p. 303) defined efficiency and illustrated the concept of efficiency as follows:

> If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ, and $\hat{\theta}_1$ is *more efficient* than $\hat{\theta}_2$ if the variance of the sampling distribution of $\hat{\theta}_1$ is less than the variance of the sampling distribution of $\hat{\theta}_2$.

FIGURE 2

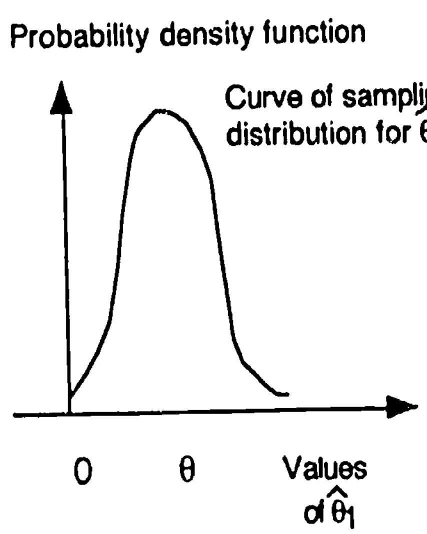
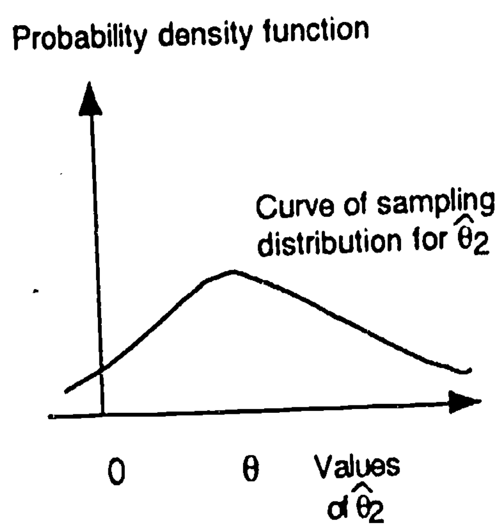


FIGURE 3



Since the variance in Figure 2 is less than the variance in Figure 3, it can be said that θ_1 is more efficient than θ_2. The relative efficiency, used in measuring efficiency, is the ratio of the variances of twc unbiased estimators.

SUFFICIENCY

Sufficiency is the last property of estimators. Harnett (1970, p. 193) defined sufficiency as:

> An estimator is said to be sufficient if it utilizes all of the information about the population parameter that is contained in the sample data.

The range is not sufficient because it only considers the highest and lowest data points. The median is not sufficient unless only ranked observations are available. The sample mean, M, is sufficient as an estimator of μ since it uses all the observed values. The variance is also sufficient since the sample mean is used in calculating the variance.

COMPUTER EXAMPLE

A BASIC computer program, written by Groeneveld (1979) and adapted by Bruce Thompson, was used to demonstrate this bias concept. The program is presented in Appendix A. A Monte-Carlo technique, defined by Danesh (1987, p. 30) as "a system of techniques which enables us to model physical systems conveniently in a computer", was used in this program. The samples were taken from a standard normal distribution (mean=0 and standard deviation=1.) The user is requested to declare a sample size "n" and the number of samples to be drawn. Since the population variance is 1.0, and variance equals the sum-of-squares/(n-1) for a sample, the expected value for the sum-of-squares(SOS) is n-1. The mean of the SOS estimates over repeated samples should equal n-1. The estimated variances should be closer to the population variance (σ^2=1) as n increases and as the number of repeated samples increases.

Examples using the program are included in Appendix B. Table 1 presents a summary of the examples presented in Appendix B. Referring to Table 1, the deviation between the expected SOS and the actual SOS tends to get smaller as either sample size or number of samples increases, as expected.

CONCLUSION

In summary, the four properties, unbiasedness, consistency, efficiency, and sufficiency,explain criteria for choosing an estimator. The properties do not specify how to find an estimator which will have some or all these properties. There are several methods, such as the method of moments, the method of maximum likelihood, and the method of least squares, which can be used to determine "good" estimators. They are not discussed in this paper and would be excellent research topics for future papers.

REFERENCES

Danesh, I. (1987). Incorporation of Monte-Carlo computer techniques into science and mathematics education. *Journal of Computers in Mathematics and Science Education,* Summer 1987, 30-36.

Eves, H. (1990). *An introduction to the history of mathematics.* Philadelphia: Saunders College Publishing.

Groeneveld, R. A. (1979). *An introduction to probability and statistics using BASIC.* New York: Marcel Dekker, Inc.

Harnett, D. L. (1970). *Introduction to statistical methods.* Reading, Massachusetts: Addison-Wesley Publishing Company.

Khazanie, R. (1990). *Elementary statistics in a world of applications.* HarperCollins Publishers, Inc.

Ott, L. (1988). *An introduction to statistical methods and data analysis.* Boston: PWS-Kent Publishing Company.

Rahman, N. A. (1968). *A course in theoretical statistics.* New York: Hafner Publishing Company.

Steel, R. G. D. & Torrie, J. H. (1960). *Principles and procedures of statistics.* New York: McGraw-Hill Book Company, Inc.

Stigler, S. M. (1986). *The history of statistics.* Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Thompson, B. (1992). Adapted BASIC computer program written by Groeneveld.

TABLE 1

Example 1	Example 2	
n=5	n=10	
k=50	k=50	
M SOS= 3.488033	M SOS= 9.637742	
d=(n-1) - M SOS	d=(n-1) - M SOS	
d_1= 0.511967	d_2= -0.637742	

Example 3	Example 4	Example 5
n=5	n=10	n=20
k=100	k=100	k=100
M SOS= 3.735797	M SOS= 8.714042	M SOS= 18.97739
d=(n-1) - M SOS	d=(n-1) - M SOS	d=(n-1) - M SOS
d_3= 0.264203	d_4= 0.285958	d_5= 0.02261

```
10 REM  PROGRAM SDMONTE.BAS
20 REM     ADAPTED BY BRUCE THOMPSON  1/92
22 OPEN "O",1,"C:SDMONTE.OUT"
24 PRINT#1," SDMONTE.OUT FROM SDMONTE.BAS":SOSTOT=0:SOS=0
26  PRINT#1,"                PROGRAM  ADAPTED  BY  BRUCE  THOMPSON
1/92":PRINT#1," "
30 DIM F(35),BT(2000):CLS
31 A$=RIGHT$(TIME$,2):S1=VAL(LEFT$(A$,1)):S2=VAL(RIGHT$(A$,1))
33 A=S1^4+S2^3:IF (A*3)<32767 THEN A=A*3
34 IF S1<5 THEN A=A*-1
36 RANDOMIZE A:PRINT#1," RANDOM NUMBER SEED WAS ";A
40 FOR I = 1 TO 35
50 LET F(I)=0
60 NEXT I
70 PRINT "WHAT IS SAMPLE";" SIZE?"
82 INPUT N:PRINT#1," SAMPLE SIZE REQUESTED WAS ";N
90 PRINT "HOW MANY";" REPITIONS?"
100 INPUT K:PRINT#1," REPITIONS OF SAMPLING REQUESTED WAS ";K
110 FOR J=1 TO K
112 PRINT#1," "
120 FOR I=1 TO N
130 GOSUB 410
140 LET S=S+Z:LET BT(I)=Z
152 PRINT#1," ";J;" ";I;" ";Z
160 NEXT I
170 LET X1=S/N
172 FOR L=1 TO N
173 LET BT(L)=BT(L)-X1:LET SOS=SOS+(BT(L)^2)
174 CLS:PRINT" ":PRINT" ":PRINT"   ";J;"  ";L
175 NEXT L
178 SOSTOT=SOS+SOSTOT
180 LET V=(SOS)/(N-1)
182 PRINT#1," SAMPLE";J:PRINT#1,"    MEAN=";X1
184 PRINT#1,"    SOS=";SOS;" N=";N;" CORRECTED V=";V
190 FOR C=1 TO 30
200 IF V>C/5 THEN 230
210   LET F(C)=F(C)+1/K
220   GOTO 260
230 IF C<30 THEN 250
240   LET F(C)=F(C)+1/K
250 NEXT C
260 LET S=0
272 LET SOS=0
280 NEXT J
290 PRINT#1," "
300 PRINT#1,"FREQUENCY ";"DISTRIBUTION OF";" VARIANCES"
310 PRINT#1," "
320 PRINT#1,"LOWER END";"   UPPER END";"   REL FREQ"
330 FOR C=1 TO 30
340 PRINT#1,(C-1)/5,C/5,F(C)
350 LET T=T+F(C)
360 IF T>=0.999999 THEN 380
370  NEXT C
380 PRINT#1," ":SOSTOT=SOSTOT/K
```

```
390 PRINT#1,"TOTAL FREQ= ";T:PRINT#1," "
392 PRINT#1,"THERE WERE";K;" SAMPLES OF SIZE";N
394 PRINT#1,"THE AVERAGE SOS SHOULD EQUAL N-1, OR";N-1
396 PRINT#1,"THE MEAN SOS OVER";K;" SAMPLES WAS";SOSTOT
400 CLOSE #1:GOTO 450
410 LET Z1=SQR(-2*LOG(RND))
420 LET Z2=6.2831853*RND
430 LET Z=Z1*COS(Z2)
440 RETURN
450 END
```

EXAMPLE 1

```
SDMONTE.OUT FROM SDMONTE.BAS
     PROGRAM ADAPTED BY BRUCE THOMPSON  1/92

RANDOM NUMBER SEED WAS -6
SAMPLE SIZE REQUESTED WAS  5
REPITIONS OF SAMPLING REQUESTED WAS  50

 1   1   .3824551
 1   2  -.1424869
 1   3   2.041772
 1   4   1.488733
 1   5   .6958976
SAMPLE 1
   MEAN= .8932742
   SOS= 3.046314   N= 5   CORRECTED V= .7615785

 2   1   1.907196
 2   2   .8472512
 2   3   1.158412
 2   4   .3580843
 2   5   .7014885
SAMPLE 2
   MEAN= .9944864
   SOS= 1.372444   N= 5   CORRECTED V= .343111

 3   1  -1.679582
 3   2   7.326465E-02
 3   3  -.1711201
 3   4  -.9935407
 3   5   .2426568
SAMPLE 3
   MEAN=-.5056642
   SOS= 2.623169   N= 5   CORRECTED V= .6557923

 4   1   .8721311
 4   2   .2723533
 4   3  -.6668983
 4   4  -.4930475
 4   5   7.062878E-02
SAMPLE 4
   MEAN= 1.103349E-02
   SOS= 1.527018   N= 5   CORRECTED V= .3817545

 5   1  -.9698443
 5   2  -1.264896
 5   3  -.3550073
 5   4  -1.034422
 5   5   .3254847
SAMPLE 5
   MEAN=-.659737
   SOS= 1.666295   N= 5   CORRECTED V= .4165736

 6   1   4.348573E-02
 6   2   .8014679
```

```
 42   3  -2.025705
 42   4    .1316303
 42   5    .4965793
SAMPLE 42
   MEAN=-.3133354
   SOS= 5.464818   N= 5   CORRECTED V= 1.366204

 43   1    .7499443
 43   2  -.2970279
 43   3   1.494941
 43   4  -1.512662
 43   5  -1.689349
SAMPLE 43
   MEAN=-.2508307
   SOS= 7.712958   N= 5   CORRECTED V= 1.928239

 44   1   1.187843E-02
 44   2  -.3291145
 44   3   .3903748
 44   4   .1354732
 44   5  -1.782711
SAMPLE 44
   MEAN=-.3148198
   SOS= 2.961703   N= 5   CORRECTED V= .7404257

 45   1  -1.365308
 45   2  -.8296563
 45   3  -1.539058
 45   4  -.1202354
 45   5  -.1267986
SAMPLE 45
   MEAN=-.7962112
   SOS= 1.781867   N= 5   CORRECTED V= .4454668

 46   1   .1714068
 46   2   .5069509
 46   3  -.125097
 46   4  -.716944
 46   5   .6539771
SAMPLE 46
   MEAN= 9.805878E-02
   SOS= 1.195646   N= 5   CORRECTED V= .2989115

 47   1  -.8491488
 47   2   .3735188
 47   3   .5075592
 47   4  -.1229607
 47   5   .4854995
SAMPLE 47
   MEAN= 7.889358E-02
   SOS= 1.337894   N= 5   CORRECTED V= .3344736

 48   1  -.3869946
 48   2   .111407
```

```
 48    3  -1.337579
 48    4  -.43302
 48    5  -3.649609E-02
SAMPLE 48
   MEAN=-.4165366
   SOS= 1.272619  N= 5  CORRECTED V= .3181548

 49    1   2.246834
 49    2   .9241474
 49    3  -1.300303
 49    4   .3597518
 49    5  -1.010265
SAMPLE 49
   MEAN= .244033
   SOS= 8.445396  N= 5  CORRECTED V= 2.111349

 50    1  -1.951918
 50    2  -.3059856
 50    3   4.825774E-02
 50    4  -.9457846
 50    5   .8500972
SAMPLE 50
   MEAN=-.4610666
   SOS= 4.460201  N= 5  CORRECTED V= 1.11505

FREQUENCY DISTRIBUTION OF VARIANCES

LOWER END   UPPER END   REL FREQ
 0            .2            .04
 .2           .4            .18
 .4           .6            .16
 .6           .8            .2
 .8           1             .08
 1            1.2           9.999999E-02
 1.2          1.4           .06
 1.4          1.6           .04
 1.6          1.8           .04
 1.8          2             .06
 2            2.2           .04

TOTAL FREQ=  .9999999

THERE WERE 50  SAMPLES OF SIZE 5
THE AVERAGE SOS SHOULD EQUAL N-1, OR 4
THE MEAN SOS OVER 50  SAMPLES WAS 3.488033
```

EXAMPLE 2

```
SDMONTE.OUT FROM SDMONTE.BAS
     PROGRAM ADAPTED BY BRUCE THOMPSON  1/92

RANDOM NUMBER SEED WAS -1797
SAMPLE SIZE REQUESTED WAS  10
REPITIONS OF SAMPLING REQUESTED WAS  50

 1   1   1.051235
 1   2  -.9002343
 1   3  -.2513088
 1   4   .0783185
 1   5  -.6785592
 1   6  -.6381989
 1   7   1.15251
 1   8   .5070424
 1   9  -.4532294
 1   10  -2.256471
SAMPLE 1
   MEAN=-.2388895
   SOS= 9.164312  N= 10  CORRECTED V= 1.018257

 2   1  -.4582848
 2   2  -.8816874
 2   3  -1.155471
 2   4   .2466978
 2   5   .6125487
 2   6   2.679835
 2   7   .1570069
 2   8   .6641071
 2   9   1.849959
 2   10  -1.624337E-02
SAMPLE 2
   MEAN= .3698468
   SOS= 12.46053  N= 10  CORRECTED V= 1.384504

 3   1   1.079786
 3   2  -.2666005
 3   3  -2.225359
 3   4   .9149314
 3   5  -1.753282
 3   6  -.6705306
 3   7   .7402433
 3   8  -.5050923
 3   9  -.9221319
 3   10   1.060137
SAMPLE 3
   MEAN=-.2547899
   SOS= 12.67806  N= 10  CORRECTED V= 1.408674

 4   1   .4536214
 4   2  -.2034944
 4   3   1.652726
 4   4   .7850095
 4   5  -.7103371
```

MEAN=-.2022671
SOS= 14.69924 N= 10 CORRECTED V= 1.633249

47	1	.4181669
47	2	.1474964
47	3	-1.929325
47	4	1.757017
47	5	1.599262
47	6	-1.337078
47	7	-.3112722
47	8	1.392713
47	9	-.4241802
47	10	1.896842

SAMPLE 47
MEAN= .3209643
SOS= 16.13574 N= 10 CORRECTED V= 1.792859

48	1	1.330566
48	2	.1593396
48	3	-.6979415
48	4	1.124701
48	5	-.8490906
48	6	-.7295691
48	7	1.155638
48	8	-1.440681
48	9	6.421478E-02
48	10	-.6832205

SAMPLE 48
MEAN=-5.660436E-02
SOS= 8.651028 N= 10 CORRECTED V= .9612252

49	1	-.9737036
49	2	-1.862909E-02
49	3	.64833
49	4	.6114015
49	5	-.9269398
49	6	-1.574742
49	7	.1073893
49	8	.9131689
49	9	-2.633728
49	10	1.323498

SAMPLE 49
MEAN=-.2523954
SOS= 13.97816 N= 10 CORRECTED V= 1.553129

50	1	.9240689
50	2	1.519857
50	3	.4949289
50	4	1.281391
50	5	-.3697424
50	6	.1431352
50	7	-.5908048
50	8	-6.102137E-02
50	9	1.623422

```
 50   10  -.3006039
 SAMPLE 50
    MEAN= .466463
    SOS= 6.110741  N= 10  CORRECTED V= .6789712

FREQUENCY DISTRIBUTION OF VARIANCES

LOWER END    UPPER END    REL FREQ
 0             .2            0
 .2            .4            .06
 .4            .6            .08
 .6            .8            .16
 .8            1             9.999999E-02
 1             1.2           .24
 1.2           1.4           .14
 1.4           1.6           .12
 1.6           1.8           9.999999E-02

TOTAL FREQ=  .9999999

THERE WERE 50  SAMPLES OF SIZE 10
THE AVERAGE SOS SHOULD EQUAL N-1, OR 9
THE MEAN SOS OVER 50  SAMPLES WAS 9.637742
```

```
SDMONTE.OUT FROM SDMONTE.BAS
     PROGRAM ADAPTED BY BRUCE THOMPSON  1/92

RANDOM NUMBER SEED WAS -27
SAMPLE SIZE REQUESTED WAS  5
REPITIONS OF SAMPLING REQUESTED WAS  100

 1   1   1.06045
 1   2  -.6737673
 1   3   .8599736
 1   4  -.0450945
 1   5  -.7684606
SAMPLE 1
   MEAN= 8.662023E-02
   SOS= 2.873121  N= 5  CORRECTED V= .7182803

 2   1  -1.124
 2   2   .8294068
 2   3  -.3992898
 2   4  -.2065119
 2   5  -.7937864
SAMPLE 2
   MEAN=-.3388363
   SOS= 2.209418  N= 5  CORRECTED V= .5523545

 3   1  -.1758775
 3   2  -.1646739
 3   3  -1.631368
 3   4  -1.158088
 3   5  -.4536565
SAMPLE 3
   MEAN=-.7167326
   SOS= 1.697854  N= 5  CORRECTED V= .4244635

 4   1  -.615754
 4   2   1.162127
 4   3  -.8984321
 4   4   .6660625
 4   5   .4664898
SAMPLE 4
   MEAN= .1560986
   SOS= 3.07629  N= 5  CORRECTED V= .7690726

 5   1   .4817174
 5   2  -.96671
 5   3  -.9364807
 5   4  -.4222362
 5   5  -1.167957
SAMPLE 5
   MEAN=-.6023333
   SOS= 1.771956  N= 5  CORRECTED V= .442989

 6   1   1.121774
 6   2  -.3654257
```

```
 96    3    7.596554E-02
 96    4   -.8280181
 96    5    .4372302
SAMPLE 96
   MEAN= 7.842743E-02
   SOS= 2.886822  N= 5  CORRECTED V= .7217056

 97    1    .4087028
 97    2   -1.486042
 97    3    1.133285
 97    4   -.4592905
 97    5   -.5068236
SAMPLE 97
   MEAN=-.1820338
   SOS= 3.96183  N= 5  CORRECTED V= .9904574

 98    1   -.4690478
 98    2   -4.930116E-02
 98    3    .3568874
 98    4   -.4764273
 98    5   -.223879
SAMPLE 98
   MEAN=-.1723536
   SOS= .478381  N= 5  CORRECTED V= .1195952

 99    1   -1.915549
 99    2    .1783297
 99    3    1.594652
 99    4    .1082491
 99    5   -1.782255
SAMPLE 99
   MEAN=-.3633146
   SOS= 8.772204  N= 5  CORRECTED V= 2.193051

 100   1   -.131384
 100   2   -.4776703
 100   3   -.1754611
 100   4   -2.606694
 100   5   -.4609728
SAMPLE 100
   MEAN=-.7704365
   SOS= 4.315707  N= 5  CORRECTED V= 1.078927
```

FREQUENCY DISTRIBUTION OF VARIANCES

LOWER END	UPPER END	REL FREQ
0	.2	.06
.2	.4	.17
.4	.6	.15
.6	.8	.17
.8	1	.13
1	1.2	.08
1.2	1.4	.04
1.4	1.6	.04

1.6	1.8	.01
1.8	2	.04
2	2.2	.04
2.2	2.4	.03
2.4	2.6	.01
2.6	2.8	.01
2.8	3	.01
3	3.2	0
3.2	3.4	0
3.4	3.6	0
3.6	3.8	0
3.8	4	.01

TOTAL FREQ= 1

THERE WERE 100 SAMPLES OF SIZE 5
THE AVERAGE SOS SHOULD EQUAL N-1, OR 4
THE MEAN SOS OVER 100 SAMPLES WAS 3.735797

EXAMPLE 4

```
SDMONTE.OUT FROM SDMONTE.BAS
     PROGRAM ADAPTED BY BRUCE THOMPSON  1/92

RANDOM NUMBER SEED WAS -849
SAMPLE SIZE REQUESTED WAS  10
REPITIONS OF SAMPLING REQUESTED WAS  100

 1   1   1.160466
 1   2  -.7973283
 1   3  -.5663562
 1   4  -.3749902
 1   5  -.7441588
 1   6   .3732632
 1   7  -.4666904
 1   8   3.724473E-02
 1   9   .3368249
 1   10  -.7284982
SAMPLE 1
   MEAN=-.1770223
   SOS= 3.686867   N= 10   CORRECTED V= .4096519

 2   1   .254492
 2   2  -1.785408
 2   3   .2480187
 2   4   .5004284
 2   5   .7799193
 2   6   .2109243
 2   7   .166352
 2   8  -.8893287
 2   9   .295921
 2   10  -.1821375
SAMPLE 2
   MEAN=-4.008183E-02
   SOS= 5.140408   N= 10   CORRECTED V= .5711564

 3   1   .1701517
 3   2   .1300232
 3   3   1.574309
 3   4   .3778581
 3   5   1.167504
 3   6   .5027012
 3   7   .5825673
 3   8  -.2600921
 3   9   .182062
 3   10  -.472255
SAMPLE 3
   MEAN= .3954829
   SOS= 3.381993   N= 10   CORRECTED V= .375777

 4   1  -.2924565
 4   2   .5006929
 4   3  -1.804697
 4   4   .8749314
 4   5   1.148781
```

```
 97   1   .8007315
 97   2  -.139967
 97   3  -.8739483
 97   4  -1.636612
 97   5   1.432342
 97   6   1.018281
 97   7  -1.181956
 97   8  -.4883819
 97   9   1.819207
 97   10   1.826599
SAMPLE 97
   MEAN= .2576295
   SOS= 14.80933   N= 10   CORRECTED V= 1.645481

 98   1   .2663514
 98   2  -.2190415
 98   3  -.9627011
 98   4  -1.614677
 98   5  -.5506749
 98   6  -1.043779
 98   7  -3.721359E-02
 98   8   .9502709
 98   9   1.157725
 98   10  -.3245938
SAMPLE 98
   MEAN=-.2378333
   SOS= 6.830055   N= 10   CORRECTED V= .758895

 99   1   6.554232E-02
 99   2   2.068519
 99   3  -.1749714
 99   4  -1.493213
 99   5  -.7910946
 99   6   .3286946
 99   7  -.3516004
 99   8   1.229989
 99   9   .1529801
 99   10  -.2458877
SAMPLE 99
   MEAN= 7.889577E-02
   SOS= 8.935349   N= 10   CORRECTED V= .9928165

 100   1  -1.596557
 100   2   .6865393
 100   3   .7121635
 100   4   1.036351
 100   5   1.284166
 100   6   .5735763
 100   7   .8916395
 100   8  -.613827
 100   9  -.4905225
 100   10  -.6363868
SAMPLE 100
```

MEAN= .1847142
SOS= 8.055815 N= 10 CORRECTED V= .8950905

FREQUENCY DISTRIBUTION OF VARIANCES

LOWER END	UPPER END	REL FREQ
0	.2	0
.2	.4	8.999999E-02
.4	.6	.19
.6	.8	.13
.8	1	.15
1	1.2	.15
1.2	1.4	8.999999E-02
1.4	1.6	.07
1.6	1.8	.08
1.8	2	.04
2	2.2	.01

TOTAL FREQ= .9999999

THERE WERE 100 SAMPLES OF SIZE 10
THE AVERAGE SOS SHOULD EQUAL N-1, OR 9
THE MEAN SOS OVER 100 SAMPLES WAS 8.714042

EXAMPLE 5

```
SDMONTE.OUT FROM SDMONTE.BAS
     PROGRAM ADAPTED BY BRUCE THOMPSON  1/92

RANDOM NUMBER SEED WAS -1536
SAMPLE SIZE REQUESTED WAS  20
REPITIONS OF SAMPLING REQUESTED WAS  100

 1   1  -.2248201
 1   2   .5194963
 1   3   2.017076E-02
 1   4   .2892476
 1   5   .5039074
 1   6  -.6579888
 1   7  -7.678485E-02
 1   8   .817639
 1   9  -.751547
 1   10  -1.217435E-03
 1   11   .2407511
 1   12   .3215078
 1   13   .3680337
 1   14   1.353352
 1   15   .3950442
 1   16   .3534379
 1   17  -1.405061
 1   18   1.386417
 1   19  -.3351373
 1   20  -.2328126
 SAMPLE 1
    MEAN= .1441818
    SOS= 8.387035  N= 20  CORRECTED V= .4414229

 2   1   4.855903E-02
 2   2   .3045103
 2   3   .8718381
 2   4  -1.468055
 2   5  -.859302
 2   6  -.244534
 2   7  -1.552927
 2   8  -.2409446
 2   9   .5322713
 2   10  -1.677375
 2   11  -.7514218
 2   12  -1.038251
 2   13  -.1949617
 2   14  -1.110046
 2   15  -.1706596
 2   16   .6479734
 2   17  -.4112441
 2   18   .4571693
 2   19  -.5458564
 2   20  -1.323593
 SAMPLE 2
    MEAN=-.4363424
    SOS= 11.357  N= 20  CORRECTED V= .5977367
```

```
97   12  -.3483107
97   13  -.1532237
97   14   6.140356E-02
97   15   .9914996
97   16   1.071523
97   17   .8948534
97   18  -.8322707
97   19   .9580903
97   20   .8942671
SAMPLE 97
   MEAN= .1711396
   SOS= 11.24199   N= 20   CORRECTED V= .5916839

98   1  -.3523787
98   2  -.9466435
98   3  -.8661198
98   4   .9511314
98   5  -.5021839
98   6  -.7183605
98   7   1.306827
98   8   .431532
98   9  -1.084639
98   10  -2.386106
98   11   .6624868
98   12   .4211462
98   13  -.3929444
98   14  -2.081142
98   15  -.1886418
98   16   .3934765
98   17  -.6623733
98   18   .5852647
98   19  -.3208738
98   20  -1.382802
SAMPLE 98
   MEAN=-.3566672
   SOS= 17.75167   N= 20   CORRECTED V= .9342986

99   1  -1.003863
99   2  -.6275403
99   3   .6826761
99   4   .1200453
99   5  -7.134871E-02
99   6   7.569096E-02
99   7  -1.01107
99   8  -.8277139
99   9   .8007606
99   10   .3926594
99   11  -.3319511
99   12   2.14052
99   13  -.3677893
99   14  -.3481696
99   15   .4060895
99   16  -.1998698
99   17   .4894969
```

```
 99   18   .5335158
 99   19   1.443584
 99   20  -.3452382
SAMPLE 99
   MEAN= 9.752419E-02
   SOS= 12.08611  N= 20  CORRECTED V= .6361111

 100   1  -1.246749
 100   2   2.044941
 100   3   .8018891
 100   4   1.941687
 100   5  -.4189042
 100   6   1.443439
 100   7   1.248126
 100   8   2.682492
 100   9  -1.521015
 100   10   1.03973
 100   11   1.197226
 100   12  -.5914441
 100   13  -.965593
 100   14   5.395073E-02
 100   15  -.3916252
 100   16   .277603
 100   17   1.33914
 100   18   .801004
 100   19  -.3931427
 100   20  -.6614072
 SAMPLE 100
   MEAN= .4340674
   SOS= 26.76395  N= 20  CORRECTED V= 1.408629

FREQUENCY DISTRIBUTION OF VARIANCES

LOWER END   UPPER END   REL FREQ
 0             .2          0
 .2            .4          0
 .4            .6          8.999999E-02
 .6            .8          .14
 .8            1           .29
 1             1.2         .27
 1.2           1.4         .11
 1.4           1.6         8.999999E-02
 1.6           1.8         0
 1.8           2           0
 2             2.2         0
 2.2           2.4         0
 2.4           2.6         .01

TOTAL FREQ=  .9999999

THERE WERE 100  SAMPLES OF SIZE 20
THE AVERAGE SOS SHOULD EQUAL N-1, OR 19
THE MEAN SOS OVER 100  SAMPLES WAS 18.97739
```